THE UNIVERSITY OF ALBERTA

AN ANALYSIS OF PERFORMANCE EXPLANATIONS OF CHILDREN'S FAILURES ON THE CLASS-INCLUSION PROBLEM

by

© PETER KASZOR

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1972

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled AN ANALYSIS OF PERFORMANCE EXPLANATIONS OF CHILDREN'S FAILURES ON THE CLASS-INCLUSION PROBLEM submitted by PETER KASZOR in partial fulfilment of the requirement for the degree of Master of Arts.

Date October 5, 1972

Abstract

Fourty-eight first graders, 48 second graders, and 48 third graders were presented a standard class-inclusion problem to determine the effects of performance factors such as <u>perceptual set</u> and <u>question misinterpretation</u> on class-inclusion difficulty. A mixed model analysis of variance design was employed. With respect to within subject factors, each subject was asked an equal number of <u>difference</u> and <u>equivalence</u> questions about the superordinate class "circles" when the subordinate classes were either numerically equal or numerically unequal. A highly significant type of question x relative size of subordinate class interaction was obtained. A significant main effect favouring the equal subordinate class condition which was expected on the basis of the perceptual set hypothesis was not obtained. With respect to between subjects variables, 12 subjects from each grade experienced one of four different methods of presentation which varied in terms of the time and extent to which concrete test stimuli were present. The main effect for presentation method was nonsignificant; the perceptual set hypothesis that totally verbal presentations should be superior to concrete presentations was generally not supported. In all cases, subjects who made class-inclusion errors were asked to recall the class-inclusion question. Very few subjects erroneously misinterpreted these questions. When misunderstood or forgotten questions were repeated to subjects a significant improvement in performance resulted although the majority of responses were still incorrect. It was concluded that the performance factors studied were not crucial determinants of class-inclusion difficulty.

Acknowledgement

My thanks to Dr. Charles Brainerd and Dr. Charles Bourassa for their help and encouragement.

Table of Contents

Introduction

Three experimental tasks have usually been used by Neo-Piagetian investigators to determine whether various cognitive structures said to develop during the concrete-operational stage of development are present or absent. These three tasks are conservation, seriation, and class-inclusion. While each task measures an ostensibly different type of cognitive ability, one of the chief assumptions of Piaget's theory is that different cognitive abilities which define a given stage develop concurrently (Flavell, 1971, 1970; Flavell & Wohlwill, 1969; Pinard & Laurendeau, 1969). There has been some research (Lovell, Mitchell, & Everett, 1962; Inhelder & Piaget, 1964) indicating that children are able to solve class-inclusion problems at around seven or eight years of age, which is roughly the same age that children begin to solve conservation and seriation problems. However, contrary to the prediction of concurrent development, recent studies have indicated that the class-inclusion problem may not be solved until late in the concrete-operatioal stage of development (e.g., nine to ten years, see Ahr & Youniss, 1970; Jennings, 1970; Kofsky, 1966, and Winer & Kronberg, 1972) at a time when seriation and some forms of conservation have been functional for two or three years (Brainerd, 1972; Bryant & Trabasso, 1971; Murray & Youniss, 1968). The most common explanations for these conflicting results are performance interpretations: the basic assumption is that a child fails the class-inclusion task because various aspects of the task adversely affect his performance. In other words, it is assumed that performance factors tend to mask the fact that the child does possess the underlying competences (i.e., appropriate cognitive structures) required to solve

class-inclusion problems.

The aim of the present study was to examine two performance hypotheses which other investigators have proposed to account for why children fail class-inclusion. One of the performance factors concerned the adverse effects which supposedly result from the perception of the stimuli used in the class-inclusion task; the other concerned the degree to which the class-inclusion question was inherently difficult. Before considering these performance factors, however, the class-inclusion task and its theoretical significance will first be described.

The Class-Inclusion Problem

According to Piaget (cf. Inhelder & Piaget, 1964) the child's ability to add classes depends on a particular cognitive structure that develops during the concrete-operational period. More specifically, this structure, referred to as Groupement I or the "primary addition of classes", is one of the nine groupement structures which Piaget postulates for this period. Normally, the presence or absence of this particular structure is assessed by means of the class-inclusion problem: the child is shown two or more numerically unequal sets of objects (i.e., six brown beads and three white beads) and he is then questioned to determine whether or not he understands that the superordinate class (beads) has a greater number of elements than either of the subordinate classes. For instance, the child might be asked whether there are more beads or more brown beads in the display. Using this method, Piaget has outlined three stages in the development of the operation of class addition. In his description he employs the following notation: B denotes the superordinate class; A

denotes the larger of the subordinate classes; A' denotes the smaller of the two subordinate classes.[1]

> At the first stage, the child is not yet capable of understanding that the B classes will always contain more elements than the A classes, the reason being that he cannot think simultaneously of the whole B and of the parts A and A', which means, logically, that he does not yet regard class B as resulting from the addition of A and A', and class A as resulting from the subtraction of A' from B. During the second stage, the child gradually comes to understand that B classes contain more elements than the A classes, but he makes this discovery intuitively . . . Finally, during the third stage, the child grasps immediately that class B is larger than class A, because he approaches the problem from the point of view of additive composition.
>
> (Piaget, 1952, pp. 163-164)

Hence, the principal Piagetian inference from failures to solve the class-inclusion problem is that the cognitive structure in question has not been acquired. Until the relationship of A + A' = B is internalized in the form of a structure, the child will continue to respond incorrectly to the class-inclusion problem.

Why Children Fail Class-Inclusion

While Piaget's research has generally involved assessing the competences (i.e., Groupement) presupposed by solutions of the class-inclusion problem, several investigators have more recently been concerned with the performance aspects of the class-inclusion task (Ahr & Youniss, 1970; Jennings, 1970; Klahr & Wallace, 1971; Kohnstamm, 1967; and Schwartz, 1970). Here, the implicit assumption is that the child's level of competence is not fairly estimated because various aspects of the class-inclusion task

---

[1] In Piaget's notation system, the union operator is designated by the arithematic sign for addition while the intersection operator is designated by the arithematic sign for multiplication.

preclude the actualization of underlying competences.

In the present study, two performance factors were studied. The first performance factor can be termed "perceptual set" or "perceptual disparity". The basic assumption here is that a subject is somehow distracted by the concrete stimuli before him, so that he tends to make a comparison between subordinate classes rather than the required comparison between a subordinate class and the superordinate class. This factor has been used to account for two types of experimental results. First, Winer and Kronberg (1972) and Wohlwill (1968) use the perceptual set argument to account for the greater ease with which children solve the class-inclusion problem under conditions of verbal presentation as opposed to pictorial (i.e., concrete) presentation. Second, the perceptual set explanation was also used by Ahr and Youniss (1970) to explain the finding that class-inclusion reasoning seems to improve as the difference (i.e., the number of items) between each subordinate class decreases.

The second performance hypothesis involves the argument that the class-inclusion question is easily misinterpreted. Several researchers (Klahr & Wallace, 1971; Kohnstamm, 1967, and Ahr & Youniss, 1970) have contended that a subject often incorrectly translates the class-inclusion question into a question involving a subordinate-subordinate class comparison, i.e., asked to compare A and B, the subject supposedly drops B and substitutes A' in the question. Given that such a mistake occurs, the subject obviously would have difficulty with the class-inclusion problem. In the sections which follow, this hypothesis as well as the perceptual set hypothesis will be considered in detail. More specifically,

each hypothesis will be examined in terms of (a) previous research, (b) rationale for its present study and (c) present hypotheses.

Perceptual Set: 1. Method of Presentation

Previous Research

Wohlwill (1968) compared the performance of five-and-six-year-old subjects who were presented with class-inclusion problems either verbally (verbal mode) or pictorially (concrete mode). In three of the four experiments conducted, the verbal mode resulted in significantly better performance. Wohlwill argued that the concrete mode made it more difficult to compare simultaneously a subordinate class with a superordinate class because of the distracting perceptual imbalance between subordinate classes A and A'. According to Wohlwill, "the perception of two contrasting subclasses, unbalanced as to number creates a strong tendency to translate a class-inclusion question into a subclass-comparison question" (Wohlwill, 1968, p. 462). This tendency to make a subordinate class comparison is presumably weakest when the items are presented in purely verbal form. By adopting this interpretation of the results, Wohlwill avoided contradicting Piaget's basic premise that a cognitive structure must first serve to assimilate or interpret concrete information before being capable of assimilating abstract (verbal) information (Piaget, 1950, Inhelder & Piaget, 1964). Wohlwill's performance argument allowed for the fact that children may have the competence to solve the class-inclusion problem but that, paradoxically, the concrete method of presentation interferes with the proper measurement of this competence.

More recent investigations of this problem have not always supported

Wohlwill's verbal facilitation effect. Jennings (1970) found that first to third graders performed significantly better when class-inclusion questions were presented concretely rather than verbally. Kindergarten subjects initially performed better ($p = .18$) in the verbal mode condition, but this trend in the data virtually disappeared when the subjects' explanations for their answers were taken into consideration. Schwartz (1970) compared the percentage of correct responses made by fifth graders in a concrete mode condition with the percentage of correct responses made by sixth graders in a verbal mode condition. The concrete mode group made a greater number of correct responses (70% versus 55%). These results clearly do not support the verbal facilitation effect and are theoretically a more plausible set of findings. However, Wohlwill has replicated his original results[1] and one very recent experiment (Winer & Kronberg, 1972, in press) does lend support to Wohlwill's findings.

Rationale

The present study attempted to resolve the issue of whether or not performance factors caused one method of presentation to be superior to other presentation methods. To do this, the competence and performance aspects of the class-inclusion task were differentiated. For instance, if the above methods of presentation were to be described in terms of "encoding" and "decoding" processes, the pictorial mode could be characterized as a "concrete encoding/concrete decoding condition", while the verbal mode could be characterized as a "verbal encoding/verbal decoding"

[1] J. Wohlwill, personal communication, May, 1972.

condition. The encoding-decoding distinction is important because the distinction aids in determining whether differences between particular presentation methods are due to performance factors, to competence factors, or to both. As Klahr and Wallace (1971) have noted, if one tends towards a performance interpretation of class-inclusion errors, Wohlwill's verbal facilitation effect should logically be due to some sort of encoding advantage. If, on the other hand, it could be shown that the advantage of Wohlwill's verbal presentation condition was exclusively a decoding advantage, it would seem reasonable to conjecture that the capacity to decode abstract quantitative information about classes is developmentally prior to the corresponding capacity to decode concrete quantitative information about classes. A decoding advantage, therefore, would suggest developmental discrepancies in these two forms of quantitative competence. Of course, to say (as in Wohlwill's case) that the verbal encoding/verbal decoding condition significantly facilitated class-inclusion problem solving does not indicate whether the determining factor is encoding alone, decoding alone, or both. More conditions obviously need to be employed.

A method of distinguishing between encoding and decoding is suggested by an experiment by Smedslund (1964). Smedslund used a presentation condition in which concrete stimuli were hidden from the subject before he made his class-inclusion judgments--in other words, it was a concrete encoding/verbal decoding type of condition. His results are interesting because approximately 50% of Smedslund's subjects correctly answered the class-inclusion questions by six years of age, much earlier than is usually reported. A second, more typical, condition where concrete stimuli were

always present was found to be slightly less effective than Smedslund's first condition. However, this second set of results may have been confounded by the fact that children always experienced the hidden stimuli condition first. Hence, while Smedslund's results were not always conclusive, his experiment is important because a concrete encoding/verbal decoding type of presentation method was used.

The following four methods of presentation were consequently studied: (a) concrete encoding/concrete decoding (CE/CD), (b) concrete encoding/verbal decoding (CE/VD), (c) verbal encoding/verbal decoding (VE/VD), and (d) verbal encoding/concrete decoding (VE/CD). Subjects in the CE/CD condition were asked to count concrete stimuli which were always present when they made their judgments. This is the typical "pictorial" or "concrete" method used by Jennings (1970), Schwartz (1971) and Wohlwill (1968). Subjects in the CE/VD condition counted concrete stimuli which were removed before they made their judgments. This procedure is similar to the one employed by Smedslund. Subjects in the VE/VD condition made judgments about a hypothetical class of elements described to them by the experimenter. Concrete stimuli were never present. This condition is similar to the "verbal" condition used by Jennings, Schwartz, and Wohlwill. Finally, subjects in the VE/CD condition were shown the concrete stimuli only when the class-inclusion questions were asked. Before these questions had been asked, the experimenter verbally described the various classes of elements.

Hypotheses

Three predictions were made. First, since Piaget assumes that concrete

reality is assimilated before abstract reality, it was predicted that the concrete decoding conditions would result in a greater number of correct responses than the verbal decoding conditions. Second, if Wohlwill's assumption about perceptual set is correct, it was predicted that the verbal encoding conditions would be superior to any set of conditions involving concrete encoding. Third, if Wohlwill's results represent a theoretical contradiction of the Piagetian position, it was predicted that the verbal decoding conditions would be superior to the concrete decoding conditions.

Perceptual Set: 2. Relative Size of Subordinate Classes and Type of Question.

Previous Research

Class-inclusion errors have also been examined by Ahr and Youniss (1970) who studied the effects of subordinate class size on class-inclusion performance. Using a concrete mode of presentation these investigators found that the greater the difference in the number of items which formed the two subordinate classes, the poorer the performance of subjects on the class-inclusion problems. This again supports the general presumption enunciated by Wohlwill that the child is distracted by the perceptual configuration before him, with there being a tendency to translate a class-inclusion question into a subordinate class comparison question as the difference between subordinate classes increased.

The effects of relative size of subordinate classes have also been studied with respect to mode of presentation. If (as Wohlwill contends) perceptual disparity between the number of elements in each subclass is

the main reason for the superiority of the verbal mode of presentation, it would then seem that verbal facilitation should be markedly reduced in situations where the subclasses are numerically equal. This prediction was supported by Jennings (1970) who found no significant differences between verbal and pictorial presentations when the subordinate classes were equal. As indicated above, Jennings was unable to replicate Wohlwill's findings when the subordinate classes were unequal; in fact, the only significant difference was one in which the concrete presentation method resulted in more correct responses in comparison to the verbal method of presentation.

Rationale

Past research has shown that Ss tend to make one of two types of judgments when presented with a class-inclusion question. Either the subject's judgment is based on a comparison of a subordinate class with a superordinate class or it is based on a comparison of the two subordinate classes. If the subject compares a subordinate class with a superordinate class he would make a conceptually correct response. In the present paper, a response based on this type of comparison will be referred to as a *conceptual* response. If, on the other hand, the subject merely compares the subordinate classes he would make a conceptually incorrect response. A response based on this type of comparison will be referred to as a *juxtaposition* response.

The degree to which a response can be assumed to be either conceptual or juxtaposition depends, in part, on the type of class-inclusion question being asked and the relative size of the subordinate classes being presented. There are two types of class-inclusion questions:

difference questions and equivalence questions. Difference questions are of the form "are there more A than B"; "are there less B than A," etc. Equivalence questions are of the form "are there the same number of B as there are A," etc. There are two general quantitative relations that one can obtain between the subordinate classes: subordinate classes are either numerically equal or numerically unequal. Both type of question and relative size of the subordinate classes must be considered in order to determine whether a response represents conceptual or juxtaposition reasoning.

Consider, for example, the case of difference questions being asked in a situation where the subordinate classes are equal (i.e., A = 5, A' = 5, and B = 10). Asking the difference question "are there more A than B" results in a conceptual response "no" if a subject correctly compares the subordinate class with a superordinate class and realizes that A is smaller than B. At the same time, if a subject incorrectly compares the two subordinate classes and observes that A is not larger than A' then the appropriate juxtaposition response is also "no." Here then is an instance where there might be responses which appear conceptually correct but which really represent juxtaposition reasoning.

Consider now the opposite situation where the subordinate classes are unequal (i.e., A = 8, A' = 2, B = 10). Asking the difference question "are there more A than B" results in a conceptual response "no". The appropriate juxtaposition response, on the other hand, is "yes". The "yes" response is appropriate because a subject who compares the two subordinate classes realizes that A > A'.

It is apparent from these examples that it is not possible to dis-

tinguish between conceptual and juxtaposition reasoning when difference questions are asked in situations where the relative size of the subordinate classes are equal. This same false agreement problem also exists when equivalence questions are asked. However, in the case of equivalence questions, the problem of false agreement exists in situations where the subordinate classes are unequal. This difference between the two types of class-inclusion questions is illustrated in Table 1.

Table 1

Format of Class-Inclusion Questions with Appropriate Conceptual and Juxtaposition Judgments

Question	Format	Equal Subsets		Unequal Subsets	
		Conceptual	Juxtaposition	Conceptual	Juxtaposition
Equivalence	Are there the same # of A as there are B	NO	YES	NO	NO
Equivalence	Are there the same # of B as there are A	NO	YES	NO	NO
Difference	Are there less B than there are A	NO	NO	NO	YES
Difference	Are there more A than there are B	NO	NO	NO	YES

When subordinate classes are equal, the use of equivalence questions allows us to distinguish conceptual and juxtaposition responses, while false agreement exists when difference questions are asked. Conversely, when the subordinate classes are unequal, difference questions allow us to distinguish conceptual and juxtaposition responses, while false agree-

ment exists when equivalence questions are asked.

This problem of false agreement is a confounding factor in the Ahr and Youniss study. Ahr and Youniss found that more conceptual judgments were made in an equal subset situation as compared to an unequal subset situation--they consequently suggested that perceptual disparity accounted for the poor performance of subjects when the subsets were unequal. Since Ahr and Youniss only employed difference questions in their study, the same trend in their results could be expected on the basis of the above analysis of the false agreement phenomenon. An examination of the difference questions used in the Ahr and Youniss study does indicate that there was a greater probability of false agreement occurring in the equal versus the unequal subset conditions. In the present study, difference *and* equivalence questions were asked in situations where the subordinate classes were numerically equal or unequal in order to control for the effects of false agreement.

Hypotheses

An interaction was predicted to occur between type of question and relative size of subordinate class. More specifically, it was predicted that approximately the same number of correct responses would be made in the equal subset condition when equivalence questions were used and in the unequal subset condition when difference questions were used. Significantly more correct responses were expected to be made in the equal subset condition when difference questions were asked and in the unequal subset condition when equivalence questions were asked.

Misinterpretation of the Class-Inclusion Question

Previous Research

Various researchers (Klahr & Wallace, 1971; Kohnstamm, 1967; Wohlwill, 1968) have suggested that the "tricky" nature of the class-inclusion question partially accounts for the difficulty of the class-inclusion problem. Both Klahr and Wallace (1971) and Kohnstamm (1967) argue that a question which asks "is there more A or more B" can sometimes be mistaken by adult subjects to be a question concerning a comparison between two subordinate classes. The same type of problem is sometimes encountered in conservation studies where subjects make a single response to questions that have a number of separate parts. Rothenberg (1969) has reviewed the literature concerning the structure of the conservation question, and suggests that single-event questions presented in several forms have an advantage over three- or two-part questions. Hence, in the case of class-inclusion questions, rather than asking "is there more B or more A," a single event question in the form "is there more B than A" might be easier to understand.

Rationale

The degree to which the class-inclusion question was misinterpreted was studied by asking subjects who made incorrect judgments to repeat the class-inclusion question. This was done to determine whether the child had changed the intent of the original question from a comparison of A and B to a comparison of A and A'. In a sense, therefore, this was a measure of whether the child properly encoded the class-inclusion question. Assuming that misinterpretations did occur, it would also be expected that there would be a reduction in class-inclusion errors if subjects were

corrected on those questions which they misunderstood. Hence, class-inclusion questions were carefully repeated when subjects misinterpreted the questions and the responses which followed were recorded.

It should also be noted that the form of the class-inclusion question used in the present study was not the typical question which asks whether "there are more A or more B." Rather, the question form was "is there more A than B" etc. This latter form does not seem to imply so strongly that the question asked involves a comparison between two separate classes or entities.

Hypotheses

It was expected that subjects who had to repeat the class-inclusion question would frequently make the erroneous A and A' comparison. This should be especially true in the concrete presentation conditions if perceptual set was a determining performance factor. It was also expected that there would be an improvement in class-inclusion performance after misunderstood questions were corrected by the experimenter.

Summary of the Purposes of the Present Experiment

The present thesis study had four empirical aims.

1. In view of the conflicting findings with respect to concrete versus verbal methods of presentation, the initial aim was to determine whether or not Wohlwill's (1968) verbal facilitation effect could be replicated.

2. Assuming that the verbal facilitation effect is replicable, the second aim of the present study was to determine whether a verbal presentation method provides mainly an <u>encoding</u> advantage or a <u>decoding</u> advantage.

As indicated above, an encoding advantage would lend more support to a performance interpretation of class-inclusion difficulty, while a decoding advantage would lend more support to a competence interpretation of class-inclusion difficulty.

3. The third aim of the present study was to re-examine the Ahr and Youniss finding that class-inclusion improved as the numerical disparity between subordinate classes was reduced. With type of question (equivalence versus difference) controlled, it was predicted that there would be a significant first-order interaction between type of question and relative size of subordinate class; it was also predicted that the relative size of the two subordinate classes would not contribute significantly to class-inclusion errors.

4. The fourth aim of the present experiment was to analysis the degree to which the class-inclusion question was misinterpreted. The experiment was designed to determine whether or not subjects translated the original class-inclusion question from a comparison of A and B to an erroneous comparison of A and A', and if so, whether there was some improvement once these misinterpretations were corrected by the experimenter.

Method

Subjects

One hundred and forty-four subjects were randomly selected from two public schools located in middle-class areas of Edmonton. Forty-eight subjects each were selected from the first, second and third grades. The subjects' mean ages were 82 months (first graders), 92 months (second graders), and 108 months (third graders). There were 20 males in the first grade sample, 27 males in the second grade sample, and 25 males in the third grade sample.

Materials

The stimuli were two sheets of white, laminated cardboard each measuring 30 x 28-cm. Ten colored circles measuring 3-cm. in diameter were printed on one side of each sheet. Eight red circles and two blue circles were printed on one sheet; the other side of the sheet had wavy lines drawn over its surface. Five red circles and five blue circles were printed on the second sheet; the other side of the sheet had horizontal drawn across its surface. The red and blue circles were grouped separately at opposite ends of each sheet.

Design

A 4 (presentation method) x 3 (age) x 2 (question presentation order) x 2 (stimulus presentation order) x 2 (type of class-inclusion question) x 2 (stimuli) mixed model analysis of variance design was employed. The first four factors were between-subjects factors. The four presentation methods were: concrete encoding/concrete decoding, concrete encoding/verbal

decoding, verbal encoding/verbal decoding, and verbal encoding/concrete decoding. The three age levels were: first grade, second grade, and third grade. The two question presentation orders were: equivalence questions-difference questions and difference questions-equivalence questions. The two stimulus presentation orders were: equal subsets (five red circles and five blue circles)-unequal subsets (eight red circles and two blue circles) and unequal subsets-equal subsets. The two remaining factors were within-subject factors. The two types of class-inclusion questions were: equivalence questions and difference questions. The two stimuli were: equal subsets and unequal subsets.

Procedure

Each subject was tested individually. The experimenter first introduced himself and asked the subject's name; he then seated himself and the subject at a 1.3 x .6 m. table. The experimenter informed the subject that they were going to play a number game.

Each subject was assigned randomly to one of the four presentation conditions. A total of 36 subjects (12 first graders, 12 second graders, and 12 third graders) were assigned to each presentation condition. Eighteen of the subjects in each condition received equivalence questions first and difference questions second, while the remaining 18 subjects received the questions in the reverse order. Eighteen of the subjects in each condition received the stimulus containing equal subsets first and the stimulus containing unequal subsets second, while the remaining 18 subjects received the opposite order. Question order and stimulus order were fully

crossed in each condition. The assessment procedure for the presentation conditions are described separately below.

Concrete encoding/concrete decoding (CE/CD). The experimenter placed one of the two stimuli in the center of the table and asked the subject, "What are all these things called?". After the subject had replied that they were all circles, the subject was instructed first to count the items in one of the subsets (red or blue circles) and then to count the items in the remaining subset. Next, the subject was asked two equivalence questions ("Are there the same number of red circles as there are circles?", "Are there the same number of circles as there are red circles?") and two difference questions ("Are there less circles than there are red circles?", "Are there more red circles than there are circles?"). Each time the subject answered one of these four questions incorrectly, the experimenter asked him to repeat the question. In the event that the subject did not repeat the question correctly, the experimenter said, "I'll repeat the question over again. Listen carefully, then answer the question yes or no", and he then carefully repeated the question. For the first stimulus, therefore, a given subject was asked at least four questions and could be asked as many as eight questions. The second stimulus was introduced next and the assessment procedure was identical to the one just described for the first stimulus.

Concrete encoding/verbal decoding (CE/VD). The procedure for this condition was the same as for the CE/CD condition except for the following change: after the subject had counted the two subsets and before he was asked any questions, the stimulus was turned over so that the subject no

longer could see the red and blue circles.

Verbal encoding/verbal decoding (VE/VD). The experimenter placed one of the two stimuli in the center of the table. In contrast to the first two conditions, the stimulus was placed face down so that the subject could not see the red and blue circles. Instead, in the equal subset stimulus condition the subject saw a sheet marked with horizontal lines; in the unequal subset stimulus condition, the subject saw a sheet marked with wavy lines. In both stimulus conditions the experimenter informed the subject that there were different-colored circles printed on the underside of the stimulus, but that all the things printed on the underside of the stimulus were circles. For the equal subsets stimulus, the subject was told that five of the circles were red and five were blue. The subject then was asked the same two equivalence questions and the same two difference questions that the subjects in the first two conditions were asked. For the unequal subsets stimulus, the subject was told that eight of the circles were red and two were blue. The subject then was asked the two difference questions and the two equivalence questions.

Verbal encoding/concrete decoding (VE/CD). The procedure for this condition was the same as for the VE/VD condition except for the following change: after the verbal description of the stimulus had been given by the experimenter and before the questions were asked, the stimulus was turned face upward so that the subject could see the red and blue circles.

Results

For purposes of analysis, all correct responses were assigned the score 1 and all incorrect responses were scored as 0. Two analyses of variance were computed. The first analysis was concerned with the judgments that the subject made following the initial presentation of each of the 8 class-inclusion questions; the analysis was not concerned with the judgments made after the experimenter had repeated a misunderstood question. For this first analysis, therefore, the range of possible scores for each subject was 0-8. The second analysis was concerned with the additional judgments made after a misunderstood question had been repeated. In this second analysis, initial judgments were analyzed only if (a) the initial judgment was correct, or (b) the judgment was incorrect but the subject repeated the class-inclusion question correctly. Otherwise, the initial judgments were not analyzed and the judgments made after the experimenter had repeated the question were analyzed. For this second analysis, the range of possible scores for each subject was also 0-8. The data employed in the second analysis provide a more valid estimate of actual class-inclusion reasoning ability than the data employed in the first analysis since the second set of data control for measurement error attributable to failures to understand class-inclusion questions. A summary table showing the appropriate means and standard deviations for the data used in the first and second analysis is respectively presented in Tables 1 and 2 of the appendix.

Unless otherwise noted, post-hoc tests of the means derived from the first and second analyses were conducted using Tukey's HSD test. When

means in an interaction table were compared, the number of treatments (K) used to obtain the q statistic was adjusted on the basis of the number of unconfounded comparisons in the table (see Cicchetti, 1972). For instance, in a 2 x 2 interaction table only four of the six possible paired comparisons are unconfounded because two comparisons are not readily interpretable. Cicchetti (1972) argues that if $K = 4$ is used to obtain the q statistic, the experimenter is penalized by being forced to accept a minimal significant difference based upon six comparisons when only four can be meaningfully accepted. The Cicchetti approximation was therefore used to adjust the K value so that in the case of a 2 x 2 interaction table, $K' = 3$ rather than four. The Tukey test is consequently more liberal because of the smaller q value that results from this adjustment.

First Analysis of Variance

A summary of the first analysis of variance appears in Table 3. There was a significant age main effect with post-hoc tests indicating that third graders made significantly more correct judgments than either first graders or second graders ($p < .01$ in both instances). An orthogonal polynominal analysis of this age trend indicated that the linear component was significant ($F = 23.51$, $df = 1/96$, $p < .01$) and that the quadratic component was not significant. Age interacted significantly with question order (see Figure 1) with first and third graders making slightly--though not significantly--more correct judgments when they received equivalence questions first. For second graders the opposite was true.

The main effect for presentation method did not attain significance.

Non-pairwise comparisons of the four presentation conditions by the Scheffé method revealed no significant differences. However, since specific predictions were made regarding methods of presentation, an "a priori" orthogonal analysis was conducted. A "t" test comparison of the verbal encoding conditions (VE/VD + VE/CD) with the concrete encoding conditions (CE/CD + CD/VD) indicated that subjects in the verbal encoding conditions made significantly more correct judgments ($t = 2.78$, $df = 96$, $p < .01$). A similar comparison of the two concrete decoding conditions (CE/CD + VE/CD) with the verbal decoding conditions (VE/VD + CE/VD) proved insignificant. The tendency for the verbal encoding conditions to be superior to the concrete encoding conditions was also demonstrated by the significant age x presentation method interaction (see Figure 2). Tests of the means for each age level indicated that second graders in the VE/CD conditions made significantly ($p < .01$) more correct judgments than subjects in the CE/CD condition. Grade 3 subjects in the CE/VD condition made significantly fewer correct judgments than subjects in the other three conditions ($p < .05$ for the CE/CD condition; $p < .01$ for the other two conditions). Only at the first grade level were the concrete encoding conditions slightly (though not significantly) superior to the verbal encoding conditions. Overall, however, it appears that the two types of encoding methods may have been virtually equivalent except for the surprisingly poor performance of third graders in the CE/VD condition. The performance of these subjects ($N = 12$) was slightly below the performance of first graders in the same condition.

Type of question was the only within-subject main effect which proved

to be significant. Difference questions ($\bar{X}$ = 2.67) led to slightly more correct judgments than equivalence questions ($\bar{X}$ = 2.48). Of the interpretable interactions two involved type of question. First and most important, there was an almost perfect interaction between type of question and stimulus. This was the interaction that was predicted earlier on the basis of an analysis of Ahr and Youniss' (1970) methodology. The interaction is presented graphically in Figure 3. Post-hoc tests indicated that equivalence questions resulted in many more correct judgments than difference questions when the unequal-subset stimulus was used ($p < .001$); however, difference questions resulted in many more correct judgments than equivalence questions when the equal-subsets stimulus was used ($p < .001$). This is precisely the type of finding that was predicted. The second interaction involving type of question was a type of question x stimulus x age interaction. Post-hoc tests revealed that the first order interaction of type of question x stimulus is less pronounced--though still significant--for third graders than for first and second graders. This is reasonable because third graders made fewer incorrect judgments than either first or second graders.

Second Analysis of Variance

A summary of the second analysis of variance appears in Table 5. This analysis included the additional judgments made after a misunderstood question had been repeated. The main effect for age was again significant, and, as before, tests indicated that third graders performed significantly better than either first graders or second graders ($p < .01$ in both instances). A significant linear trend was demonstrated ($F = 17.95$, $df = 96$, $p < .01$)

but once again the quadratic component was insignificant ($F < 1$). The interactions of age with question order noted in the first analysis of variance did not emerge in this analysis. As in the first analysis, however, age was observed to interact significantly with presentation method. Tests indicated that first and second graders found the four presentation methods of equivalent difficulty, but that third graders found the CE/VD method significantly more difficult than the other three methods ($p < .01$ in the case of the VE/VD method: $p < .05$ in the other two cases).

The within-subject effect of type of question was again significant with difference questions being slightly easier than equivalence questions. The predicted interaction of type of question x stimulus was again very large. This interaction is presented graphically in Figure 4. The second-order interaction of type of question x stimulus x age noted in the first analysis did not emerge in this analysis. However, there were other interactions that were significant using the second analysis which were not significant using the first analysis. First, type of question interacted significantly with age (see Figure 5). Post-hoc tests indicated that third graders found equivalence questions significantly ($p < .01$) easier than either first or second graders. Second, type of question interacted significantly with question order (see Figure 6). Inspection of the means suggested that this interaction can be viewed as a "warm-up" effect of sorts: The second question in each ordering--regardless of the specific type of question--tended to be easier than the first question. The adjusted Tukey test revealed that difference questions resulted in

significantly more correct responses ($\underline{p}$ < .01) than did the preceding equivalence questions. Third, type of question interacted significantly with stimulus presentation order (see Figure 7). Difference questions presented in situations where the unequal subset-equal subset order was employed resulted in significantly more correct responses than did equivalence questions asked in this condition.

Of the three remaining significant F ratios in Table 5, there are two first order interactions (stimulus x question order, and stimulus x stimulus presentation order) and a third order interaction (age x presentation method x stimulus x type of question). Only the first-order interactions will be considered. Concerning the interaction of stimulus with question order (see Figure 8) an inspection of the means indicates that this is also a warm-up effect of sorts: slightly more correct responses were made in conditions where the false agreement questions came last. This was especially true in unequal subset condition; here, difference questions resulted in significantly more correct responses when they were preceded by equivalence questions (p <.01). Concerning the interaction of stimulus with stimulus presentation order (see Figure 8), it appears that a reverse warm-up effect of sorts occurred since more correct judgments were made in the first subset condition presented than in the second. Post-hoc tests demonstrated the number of correct responses made in the unequal subset condition were significantly greater than the number made in the equal subset condition when the unequal-equal subset order was employed.

Repetition of the Class-Inclusion Question

The mean number of incorrect responses made before and after class-inclusion questions were repeated are shown in Table 10. Correlated "t" tests of the means at each grade level indicated that there was a significant decrease in the number of incorrect responses made after the question was repeated. It is therefore possible that repeating the class-inclusion question helped the subject to understand the question better, and hence led to improved performance. However, this improvement was quite small. For instance, after the experimenter repeated the questions which were misinterpreted, 85% of the responses of the first graders, 84% of the responses of the second graders, and 65% of the responses of the third graders were still incorrect (see Table 11).

As might be expected, it was not necessary to repeat as many questions to the older subjects. Tables 12 and 13 indicate the number of subjects at each age level who required one correction or less (Table 12) or two corrections or less (Table 13). Chi-square analysis indicated that there were significant differences between the table values (Table 12, x^2 = 12.38, df = 2, $p < .01$; Table 13, x^2 = 9.76, df = 2, $p < .01$) with third graders clearly requiring less corrections in both instances. Looked at in another way, this usually meant that older subjects made fewer recall errors than younger subjects. While almost all the subjects at each age level were asked to recall one or more class-inclusion questions, only 20 third graders recalled one or more class-inclusion questions wrong as compared to 36 second graders and 43 first graders (see Table 14). Chi-square tests using Yates' correction indicated that there were significantly

less recall errors made by third graders as compared to second graders ($x^2 = 9.25$, df = 1, $p < .01$) or as compared to first graders ($x^2 = 16.60$, df = 1, $p < .01$).

Finally, very few subjects (9% overall) incorrectly compared the two subordinate classes when they were asked to repeat the class-inclusion question (see Table 15). Rather, subjects usually encoded the question correctly or could not remember the question at all. Older subjects were more likely to encode the question correctly than to forget the question, i.e., 67% of the third graders correctly repeated the class-inclusion question versus the 27% that forgot.

Discussion

Generally speaking, the performance factors which were studied did not seem to affect strongly the degree to which subjects correctly answer class-inclusion questions. This was especially true with regard to the proposed influence of perceptual set on class-inclusion reasoning. The results pertaining to the problem of misinterpretation of the class-inclusion question were less clear, but strong support for any type of performance argument generally was lacking.

Method of Presentation

Although several predictions were made with respect to method of presentation, only one comparison between the various methods approached statistical significance. This was the tendency for the two verbal encoding conditions to be easier than the two concrete encoding conditions. The performance of second and third graders accounted for this trend since first graders performed slightly better in the concrete encoding conditions. However, the difference between the encoding conditions was obtained only when the initial judgments made by each subject was considered. When the additional (i.e., second) judgments made by each subject were taken into account, this difference virtually disappeared. Furthermore, neither analysis of variance indicated that there were significant differences between the verbal decoding and the concrete decoding conditions.

The above results therefore do not lend convincing support to Wohlwill's argument that perceptual disparity adversely affects class-inclusion reasoning. It is true that the difference obtained between the verbal and concrete methods of presentation involved the encoding conditions.

As previously pointed out in the introduction, a discrepancy between verbal and concrete encoding probably represents a performance rather than a competence difference. However, two arguments can be made against this being strong support for Wohlwill's perceptual set hypothesis. First, this result was obtained by comparing the two verbal encoding conditions with the two concrete encoding conditions. All of the previous studies (i.e., Jennings, 1970; Schwartz, 1970; Winer & Kronberg, 1972; Wohlwill, 1968) which have dealt with method of presentation have only assessed differences between the two conditions which have been designated CE/CD and VE/VD in the present study. No significant differences were obtained between the CE/CD and VE/VD conditions using either the first or the second set of data. Viewed in this context, Wohlwill's verbal facilitation effect was not replicated in the present experiment.

Second, the verbal encoding - concrete encoding difference was not obtained when the second (additional) judgments of the subjects were analyzed. Misunderstood questions were repeated to subjects in the hope of avoiding any possible error due to question misinterpretation. As it will be pointed out later, it would be difficult to say unequivally that this second set of data was therefore "better" than the first group of data, since a control condition was not used when the additional judgments were obtained. Suffice it to say, however, that the difference between the two types of encoding conditions were virtually nonexistent once subjects had a chance to make judgments about questions they misunderstood. Considering these present results and the conflicting results previously obtained by different researchers, it would seem that, at best, the

verbal facilitation effect is a weak effect. With reference to method of presentation, therefore, it might also be argued that perceptual set does not greatly effect class-inclusion difficulty.

The Piagetian assumption that the capacity to understand concrete information precedes the capacity to understand verbal information was not supported in the present study. Both Jennings (1970) and Schwartz (1970) had previously demonstrated that the concrete (i.e., CE/CD) method of presentation resulted in a greater number of correct responses being made than the number made when presentation conditions were totally verbal. This would be the type of result expected on the basis of Piaget's theory. In the present study this result was not obtained, even though some of the procedures used (i.e., the CE/CD and VE/CD conditions) were quite similar to the procedures other investigators have termed "concrete" and "verbal." The failure to replicate Jennings' and Schwartz's results and, for that matter, Wohlwill's findings seems to imply that the concrete-verbal distinction is a rather tenuous one, at least as it applies to the class-inclusion problem. Of course, in the strict sense of the word, the "concrete" presentation methods are not totally "concrete"--class-inclusion questions are, by necessity, verbal. Perhaps it is more appropriate to say that one condition is "more verbal" than another than to assume that some sharp division exists between each presentation method in terms of the type of information supposedly presented to the subject.

The failure to find any real differences between the various presentation methods could also be interpreted as indicating that Piaget's assump-

tions about the concrete-verbal sequence of development are somewhat erroneous. For example, the idea that concrete reality must be assimilated before abstract reality may not be true on all occassions. Class addition might represent an instance where subjects can begin to understand both forms of information at approximately the same age. In order to substantiate this interpretation, more data is obviously required.

Type of Question and Relative Size of Subordinate Class

Perceptual set was also the explanation suggested by Ahr and Youniss (1970) to account for the difficulty of the class-inclusion task in situations where the subsets were unequal. This argument definitely was not supported by the present data. Both analyses indicated that the stimulus-size main effect was nonsignificant. Instead, the number of correct responses depended upon the type of question asked when the subsets were either clearly equal or unequal. When the subordinate classes were unequal, equivalence questions resulted in mostly correct judgments, while difference questions resulted in mostly incorrect judgments. When the subordinate classes were equal, equivalence questions resulted in mostly incorrect judgments, while difference questions resulted in mostly correct judgments. It should also be noted that the two analyses included the condition designated "verbal encoding/verbal decoding." Ahr and Youniss' argument was in specific reference to situations where the stimuli were concretely presented. An inspection of the means from the two sets of data makes it apparent that the inclusion of a totally verbal condition does not detract from the basic findings described above. In both

analyses a significant type of question x stimulus x grade x method of presentation interaction was obtained. What basically happened was this: younger subjects usually made more correct responses after difference questions than after equivalence questions in situations that lacked false agreement; older subjects were less likely to follow this pattern of responding. For instance, third graders in the VE/VD condition clearly made more correct responses following equivalence questions than they did following difference questions.

False agreement therefore seems to be a valid explanation for why it sometimes happens that more correct responses to class-inclusion questions occur when subordinate classes are equal. In all likelihood, if this form of error is eliminated, researchers (i.e. like Ahr and Youniss) would find that the class-inclusion problem is solved at an even later age than previously thought. The present data consequently have a twofold implication: first, the data show that perceptual disparity does not account for why children fail class-inclusion, and second, the data imply that the avoidance of false agreement questions would increase the probability that the ability to solve class-inclusion problems would occur after the ability to solve most other concrete-operational problems.

A result which was not predicted was the finding that difference questions resulted in a greater number of correct responses than did equivalence questions. Like many of the interactions involving order of stimulus presentation and order of type of question, this was a small effect compared to the very significant type of question x stimulus size interaction or even the main effect due to age. As Table 1 of the

Introduction indicates, both difference and equivalence questions had the same format except for the use of the terms "same", "more", or "less". No test was conducted to measure the subjects' understanding of these terms. It is possible that children who did not completely understand words like "more" or "less" might have guessed when the class-inclusion question was presented, thereby increasing their chances of making a correct response. Such guessing would be more likely to occur among the younger subjects who were responding below chance on most occassions. This assumption is partially supported by a significant age level x type of question interaction obtained using the second analysis of variance. This interaction demonstrates that there were more correct responses to difference questions amongst first and second graders; third graders made slightly more correct responses following equivalence questions. Overall, however, the difference between the two types of questions was quite small. Since this difference does not seem to have any inherently interesting theoretical implications, any further interpretation seems unwarranted.

Misinterpretation

The second performance explanation--misinterpretation of the class-inclusion question--was investigated in two ways. First, the number of incorrect responses made before and after the experimenter repeated the class-inclusion question were compared. Correlated "t" tests of the means at each grade level indicated that there was a significant decrease in the number of incorrect responses after the question was repeated. This might indicate that subjects were better able to understand the class-inclusion question once it was repeated. There are two problems with this

interpretation. First, it should be remembered that the majority of second responses were still incorrect even though there was an improvement after misunderstood questions were repeated. Second, some degree of hesitation also seems warranted because a control condition was not included to assess whether the mere repetition of the class-inclusion question would cause a subject to change his response. As the class-inclusion question was only repeated if the subject initially made an incorrect response, the possibility of this type of error exists.

The second method used to investigate question misinterpretation involved an analysis of the answers given by subjects who were asked to recall the class-inclusion question. The number of answers where the A and B classes were correctly remembered were compared to the number of answers where the A and A' classes were incorrectly mentioned. Both Klahr and Wallace (1970) and Kohnstamm (1967) have argued that the class-inclusion question is so tricky that incorrect responses can result from a subject thinking that he is being asked to make a comparison between subordinate classes rather than between a subordinate class and a superordinate class. Theoretically, of course, the subject makes an erroneous comparison between subordinate classes because of his inability to simultaneously think of the subordinate classes and the superordinate class. At the theoretical level a misunderstanding of the class-inclusion question is not necessarily implied--the subject understands the question, but he simply cannot do what the question asks. The data (Table 15) tend to support the theoretical (i.e., competence) argument. Almost all the subjects either correctly repeated the class-inclusion question <u>or</u> could not remember the question at all. It is conceivable that before "forgetting"

the question, subjects incorrectly interpreted the question to be a comparison of A and A'. However, a developmental trend with age was clear--subjects were more and more likely to repeat the class-inclusion question correctly, and less likely to forget it. The percentage of responses of the A and A' type did not change with age. Thus, this second group of analyses indicates that incorrect class-inclusion judgments were not, in the main, a result of question misinterpretation. Being able to repeat the question correctly did not insure a correct judgment response.

Two further points can be made with regard to the effects of question misinterpretation. First, there was no tendency for subjects to misinterpret more questions when the concrete presentation methods versus the totally verbal presentation method were used. This outcome may have been expected on the basis of the perceptual disparity argument. Second, changing the format of the class-inclusion question to a single event question requiring a "yes" or a "no" answer did not noticeably reduce the age level at which the class-inclusion problem is usually solved. Under conditions where class-inclusion questions were repeated when a subject made an incorrect response, nine-year-old subjects were able to answer about six of the eight class-inclusion questions correctly. Of course, four of the six corrects involved a false agreement question. Since the performance factors which were investigated did not seem to effect class-inclusion reasoning to any great extent, we are still left with the problem of why children fail class-inclusion. It would appear from the present data that class-inclusion failure is due to competence rather than performance factors. What is needed now are further studies

to definitely establish the probable sequential relationship between the competences underlying class-inclusion, on the one hand, and the competences underlying other concrete-operation concepts on the other.

References

Ahr, P.R., and Youniss, J. Reasons for failure on the class-inclusion problem. Child Development, 1970, 41, 131-143.

Brainerd, C.J. Order of acquisition of transitivity, conservation and class-inclusion of length and weight. Developmental Psychology, 1972, in press.

Bryant, P.E., & Trabasso, T. Transitive inferences and memory in young children. Nature, 1971, 232, 456-458.

Cicchetti, D.C. Extension of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. Psychological Bulletin, 1972, 77, 405-408.

Flavell, J.H., and Wohlwill, J.F. Formal and functional aspects of cognitive development. In D. Elkind and J.H. Flavell (Eds.), Studies in cognitive development: Essays in honor of Jean Piaget. New York: Oxford University Press, 1969, pp. 69-120.

Flavell, J.H. An analysis of cognitive-developmental sequences. Division 7. Presidential address presented at the meeting of the American Psychological Association, Miami Beach, September 1970.

Flavell, J.H. Stage-related properties of cognitive development. Cognitive Psychology, 1971, 2, 421-453.

Inhelder, B., and Piaget, J. The early growth of logic in the child. New York: Harper and Row, 1964.

Jennings, J.R. The effect of verbal and pictorial presentation on class-inclusion competence and performance. Psychonomic Science, 1970, 20,357-358.

Klahr, D., and Wallace, J.G. Class inclusion processes. Paper presented at the 7th Annual Carnegie Symposium on Cognition, May 1971.

Kofsky, Ellin. A scalogram study of classificatory development. Child Development, 1966, 37, 191-204.

Kohnstamm, G.A. Piaget's analysis of class inclusion: Right or wrong? The Hague: Mouton, 1967.

Lovell, K., Mitchell, B., and Everett, R. An experimental study of the growth of some logical structures. British Journal of Psychology, 1962, 53, 175-188.

Murray, J.P., & Youniss, J. Achievement of inferential transitivity and its relation to serial ordering. Child Development, 1968, 39, 1259-1268.

Piaget, J. The psychology of intelligence. New York: Harcourt, Brace, 1950.

Piaget, J. The child's conception of number. New York: W.W. Norton & Company, 1964 (first English edition, 1952).

Pinard, A., and Laurendeau, M. "Stage" in Piaget's cognitive developmental theory: Exegesis of a concept. In D. Elkind and J.H. Flavell (Eds.), Studies in cognitive development: essays in honor of Jean Piaget. New York: Oxford University Press, 1969, pp. 121-170.

Rothenberg, B.A. Conservation of number among four- and five-year-old children: Some methodological considerations. Child Development, 1969, 40, 383-406.

Schwartz, C.R. Developmental aspects of class inclusion. Ph.D. dissertation, University of California (Berkley), 1970.

Smedslund, J. Concrete reasoning: A study of intellectual development. Monographs of the Society for Research in Child Development, 1964, 29, Serial No. 93.

Winer, G.A., and Kronberg, D.D. Responses from children in grades kindergarten through six to class-inclusion items presented verbally and pictorially. In press, 1972.

Wohlwill, J.F. Response to class-inclusion questions for verbally and pictorially presented items. _Child Development_, 1968, _39_, 449-465.

Appendix

Table 1

MEANS AND STANDARD DEVIATIONS OF DATA
USED FOR FIRST ANALYSIS OF VARIANCE

Grade	Unequal Subclass		Equal Subclass	
	Equivalence	Difference	Equivalence	Difference
	Verbal Encoding/Verbal Decoding			
Grade 1	1.58 (.80)	.67 (.79)	.58 (.67)	2.00 (0)
Grade 2	1.75 (.62)	.75 (.75)	.33 (.79)	2.00 (0)
Grade 3	2.00 (0)	1.33 (.89)	1.17 (.94)	1.92 (.29)
	Verbal Encoding/Concrete Decoding			
Grade 1	2.00 (0)	.17 (.39)	.08 (.28)	2.00 (0)
Grade 2	1.92 (.29)	1.08 (.90)	.67 (.89)	2.00 (0)
Grade 3	2.00 (0)	1.08 (.60)	1.42 (.90)	2.00 (0)
	Concrete Encoding/Concrete Decoding			
Grade 1	1.92 (.29)	.42 (.67)	.50 (.80)	1.92 (.29)
Grade 2	1.92 (.29)	.33 (.49)	.08 (.29)	1.92 (.29)
Grade 3	2.00 (0)	1.00 (.85)	.83 (.79)	2.00 (0)
	Concrete Encoding/Verbal Decoding			
Grade 1	1.75 (.45)	.92 (.79)	.50 (.52)	1.75 (.62)
Grade 2	1.92 (.29)	.58 (.67)	.67 (.89)	1.83 (.15)
Grade 3	2.00 (0)	.42 (.69)	.17 (.58)	2.00 (0)

*Each cell based on 12 observations; maximum score is 2.

Standard deviations in brackets.

Table 2

MEANS AND STANDARD DEVIATIONS OF DATA USED FOR SECOND ANALYSIS OF VARIANCE

Grade	Unequal Subclass		Equal Subclass	
	Equivalence	Difference	Equivalence	Difference
Verbal Encoding/Verbal Decoding				
Grade 1	1.58 (.80)	.75 (.75)	.67 (.79)	2.00 (0)
Grade 2	1.75 (.62)	.83 (.72)	.50 (.80)	2.00 (0)
Grade 3	2.00 (0)	1.33 (.89)	1.44 (.89)	1.92 (.29)
Verbal Encoding/Concrete Decoding				
Grade 1	2.00 (0)	.42 (.52)	.08 (.29)	2.00 (0)
Grade 2	1.92 (.29)	1.25 (.87)	.67 (.89)	2.00 (0)
Grade 3	2.00 (0)	1.08 (.67)	1.42 (.90)	2.00 (0)
Concrete Encoding/Concrete Decoding				
Grade 1	1.92 (.29)	.83 (.83)	.50 (.80)	1.92 (.29)
Grade 2	1.92 (.29)	.67 (.89)	.08 (.29)	1.92 (.29)
Grade 3	2.00 (0)	1.17 (.84)	1.17 (.94)	2.00 (0)
Concrete Encoding/Verbal Decoding				
Grade 1	1.75 (.45)	.92 (.79)	.92 (.90)	1.75 (.62)
Grade 2	1.92 (.29)	.83 (.83)	.83 (.90)	1.83 (.15)
Grade 3	2.00 (0)	.42 (.67)	.50 (.90)	2.00 (0)

Table 3

SUMMARY TABLE: FIRST ANALYSIS OF VARIANCE

Source	df	MS	F
A (Grade)	2	4.36	13.00**
B (Mode)	3	0.87	2.62***
C (Question Order)	1	0.69	2.07
D (Stimulus Order)	1	1.00	2.98
AB	6	1.38	4.12**
AC	2	1.23	3.67*
AD	2	0.33	0.98
BC	3	0.14	0.43
BD	3	0.14	0.43
CD	1	0.007	0.02
ABC	6	0.21	0.64
ABD	6	0.51	1.51
ACD	2	0.38	1.12
BCD	3	0.51	1.53
ABCD	6	0.51	1.53
Error: Subj. within	96	0.34	
S (Question Type)	1	1.36	5.98*
AS	2	0.23	1.02
BS	3	0.21	0.92
CS	1	0.34	1.50
DS	1	0.34	1.50
ABS	6	0.31	1.35
ACS	2	0.17	0.74
ADS	2	0.31	1.38
BCS	3	0.07	0.29
BDS	3	0.06	0.25
CDS	1	0.11	0.48
ABCS	6	0.50	2.20*
ABDS	6	0.27	1.21
ACDS	2	0.23	1.02
BCDS	3	0.06	0.26
ABCDS	6	0.24	1.07
S x subj. within	96		
T (Stimulus Size)	1	0.34	1.37
AT	2	0.15	0.34
BT	3	0.08	0.31
CT	1	0.25	1.01
DT	1	0.44	1.79
ABT	6	0.26	1.04
ACT	2	0.33	1.32
ADT	2	0.69	2.77
BCT	3	0.20	0.81
BDT	3	0.14	0.58
CDT	1	0.17	0.70
ABCT	6	0.26	1.03
ABDT	6	0.20	0.81
ACDT	2	0.11	0.43
BCDT	3	0.14	0.57
ABCDT	6	0.59	2.39*
T x subj. within	96	0.25	
TS	1	230.03	589.72**
ATS	2	1.48	3.80*
BTS	3	0.80	2.05
CTS	1	0.06	0.16
DTS	1	0.84	2.15
ABTS	6	2.12	5.44**
ACTS	2	0.91	2.34
ADTS	2	0.21	0.54
BCTS	3	0.22	0.55
BDTS	3	0.26	0.67
CDTS	1	0.44	1.14
ABCTS	6	0.20	0.50
ABDTS	6	0.68	1.74
ACDTS	2	0.23	0.59
BCDTS	3	0.23	0.58
ABCDTS	6	0.66	1.69
TS x subj. within	96		

* $p < .05$

** $p < .01$

*** $p < .10$

Table 4

MEANS AND STANDARD DEVIATIONS FOR THE

AGE x QUESTION ORDER INTERACTION

Grade	Question Order	
	Difference-Equivalence	Equivalence-Difference
1	4.33 (1.06)	5.04 (1.12)
2	5.17 (1.46)	4.71 (1.05)
3	5.54 (1.32)	6.13 (1.51)

Table 5

SUMMARY TABLE: SECOND ANALYSIS OF VARIANCE

Source	df	MS	F
A (Grade)	2	4.07	9.99**
B (Mode)	3	0.31	0.76
C (Question Order)	1	0.25	0.61
D (Stimulus Order)	1	0.84	2.06
AB	6	1.43	3.50**
AC	2	0.82	2.00
AD	2	0.30	0.74
BC	3	0.10	0.23
BD	3	0.45	1.10
CD	1	0.08	0.02
ABC	6	0.36	0.88
ABD	6	0.66	1.62
ACD	2	0.39	0.95
BCD	3	0.56	1.37
ABCD	6	0.59	1.45
Error: subj. within	96	0.41	
S (Question Type)	1	1.36	5.06*
AS	2	1.11	4.11*
BS	3	0.26	0.98
CS	1	1.78	6.61*
DS	1	1.17	4.36*
ABS	6	0.15	0.55
ACS	2	0.19	0.70
ADS	2	0.13	0.47
BCS	3	0.18	0.67
BDS	3	0.08	0.28
CDS	1	0.56	2.09
ABCS	6	0.30	1.12
ABDS	6	0.21	0.78
ACDS	2	0.54	1.99
BCDS	3	0.08	0.28
ABCDS	6	0.11	0.39
S x subj. within	96	0.27	
T (Stimulus Size)	1	0.34	1.22
AT	2	0.52	1.87
BT	3	0.27	0.97
CT	1	1.56	5.59*
DT	1	1.36	4.87*
ABT	6	0.16	0.58
ACT	2	0.41	1.47
ADT	2	0.38	1.35
BCT	3	0.16	0.57
BDT	3	0.18	0.65
CDT	1	0.69	2.48
ABCT	6	0.23	0.83
ABDT	6	0.17	0.61
ACDT	2	0.56	2.02
BCDT	3	0.07	0.25
ABCDT	6	0.34	1.22
T x subj. within	96	0.28	
TS	1	180.00	391.79**
ATS	2	1.32	2.88
BTS	3	0.44	0.95
CTS	1	0.005	0.02
DTS	1	0.69	1.52
ABTS	6	2.12	4.60**
ACTS	2	0.43	0.93
ADTS	2	0.43	0.93
BCTS	3	0.18	0.39
BDTS	3	0.61	1.32
CDTS	1	0.25	0.54
ABCTS	6	0.22	0.49
ABDTS	6	0.97	2.10
ACDTS	2	0.32	0.69
BCDTS	3	0.24	0.51
ABCDTS	6	0.61	1.33
TS x subj. within	96		

* $p < .05$

** $p < .01$

Table 6

MEANS AND STANDARD DEVIATIONS FOR THE TYPE OF QUESTION x QUESTION ORDER INTERACTION

Order	Type of Question	
	Equivalence	Difference
Difference-Equivalence	2.69 (1.00)	2.67 (0.79)
Equivalence-Difference	2.56 (0.92)	2.97 (0.79)

Table 7

MEANS AND STANDARD DEVIATIONS FOR THE TYPE OF QUESTION x STIMULUS PRESENTATION ORDER INTERACTION

Order	Type of Question	
	Equivalence	Difference
Equal-Unequal	2.64 (0.98)	2.65 (0.77)
Unequal-Equal	2.61 (0.94)	2.99 (0.80)

Table 8

MEANS AND STANDARD DEVIATIONS FOR THE STIMULUS x QUESTION ORDER INTERACTION

Order	Stimulus	
	Unequal	Equal
Difference-Equivalence	2.63 (0.92)	2.74 (0.97)
Equivalence-Difference	2.92 (0.80)	2.61 (0.87)

Table 9

MEANS AND STANDARD DEVIATIONS FOR THE STIMULUS x STIMULUS PRESENTATION ORDER

Order	Stimulus	
	Unequal	Equal
Equal-Unequal	2.60 (0.84)	2.70 (0.93)
Unequal-Equal	2.94 (0.86)	2.65 (0.92)

Table 10

MEAN NUMBER OF INCORRECT RESPONSES BEFORE AND AFTER CLASS-INCLUSION QUESTIONS WERE REPEATED

Before	After	t Values
Grade 1		
2.30 (1.06)	1.95 (1.15)	$t = 4.38$, $df = 42$, $p < .001$
Grade 2		
2.47 (0.93)	2.08 (1.14)	$t = 3.39$, $df = 35$, $p < .002$
Grade 3		
1.85 (0.85)	1.20 (0.93)	$t = 3.32$, $df = 19$, $p < .004$

Table 11

NUMBER OF CORRECT AND INCORRECT RESPONSES MADE AFTER CLASS-INCLUSION QUESTIONS WERE REPEATED

Grade	N	Response	
		Correct	Incorrect
1	43	15 (15%)	84 (85%)
2	36	14 (16%)	75 (84%)
3	20	13 (35%)	24 (65%)

Table 12

NUMBER OF SUBJECTS AT EACH GRADE LEVEL PERFECT OR CORRECTED ONCE VERSUS THE NUMBER CORRECTED MORE THAN ONCE

Grade	Required Corrections	
	1 or less	More than 1
1	5	43
2	7	41
3	18	30

Table 13

NUMBER OF SUBJECTS AT EACH GRADE LEVEL REQUIRING 2 CORRECTIONS OR LESS VERSUS THE NUMBER REQUIRING MORE THAN 2 CORRECTIONS

Grade	Required Corrections	
	2 or less	More than 2
1	14	34
2	13	35
3	27	21

Table 14

NUMBER OF SUBJECTS AT EACH GRADE LEVEL MAKING NO RECALL ERRORS VERSUS THE NUMBER MAKING ONE OR MORE RECALL ERROR

Grade	Recall Errors	
	None	One or more
1	4	43
2	7	36
3	20	20

Table 15

FORM OF CLASS-INCLUSION QUESTIONS REPEATED BY SUBJECTS CONSIDERED OVER AGE LEVEL AND METHOD OF PRESENTATION

Form of Answer	Presentation Method			
	VE/VD	VE/CD	CE/CD	CE/CD
Grade 1				
A and B	10 (31%)	10 (27%)	11 (37%)	14 (38%)
A and A'	4 (12.5%)	4 (11%)	2 (6%)	0 (0)
A' and B	4 (12.5%)	2 (5%)	3 (10%)	2 (6%)
Forget	14 (44%)	21 (57%)	14 (47%)	21 (58%)
Grade 2				
A and B	17 (52%)	7 (27%)	15 (35%)	10 (29%)
A and A'	4 (12%)	5 (19%)	5 (12%)	2 (6%)
A' and B	1 (3%)	1 (4%)	3 (7%)	2 (6%)
Forget	11 (33%)	13 (50%)	20 (47%)	20 (59%)
Grade 3				
A and B	12 (67%)	15 (83%)	15 (56%)	24 (60%)
A and A'	1 (5%)	1 (6%)	1 (4%)	4 (10%)
A' and B	0 (0%)	0 (0%)	1 (4%)	0 (0)
Forget	5 (28%)	2 (11%)	10 (37%)	12 (30%)

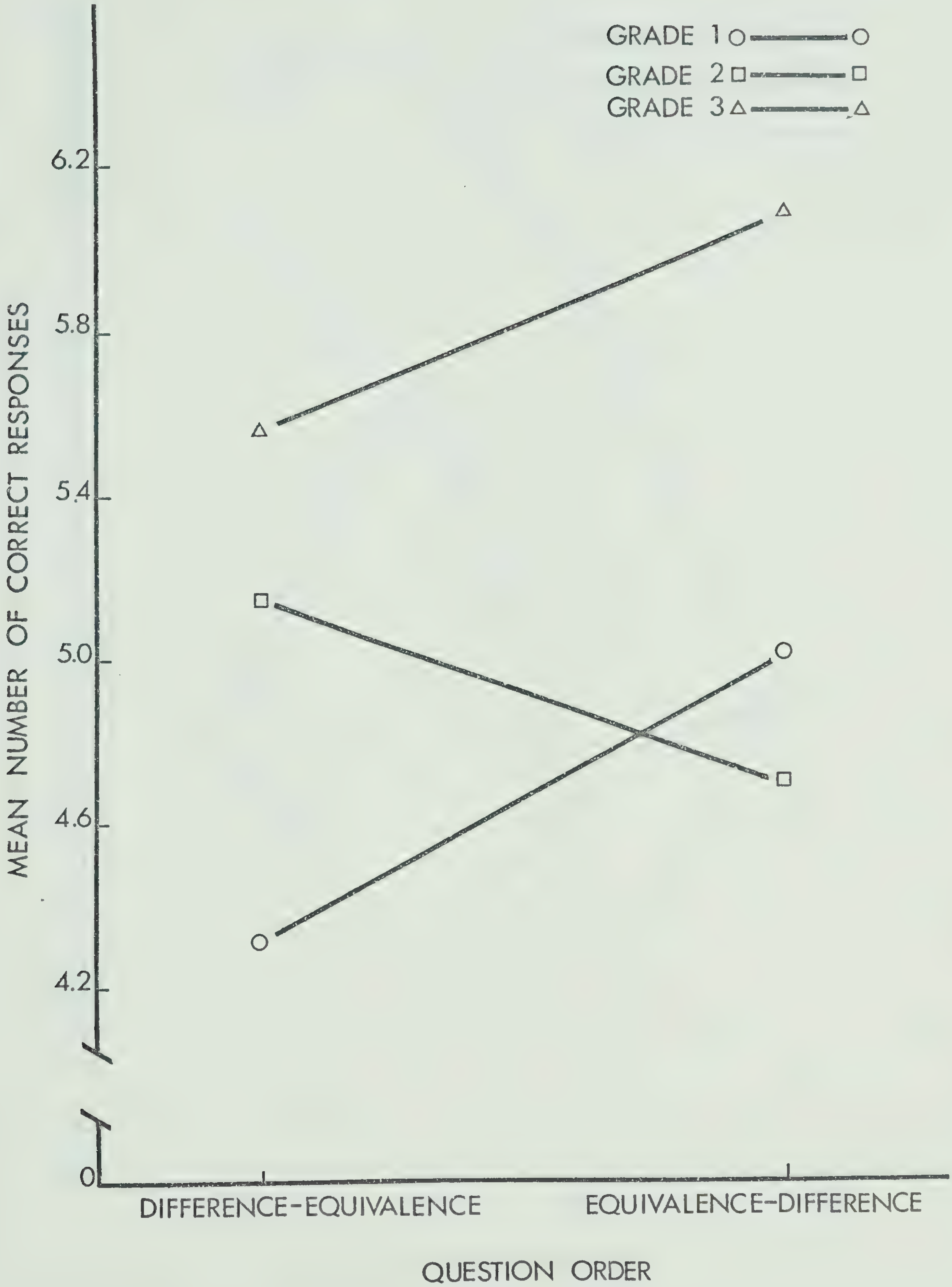

Figure 1: Graph of age x question order interaction based on means presented in Table 4.

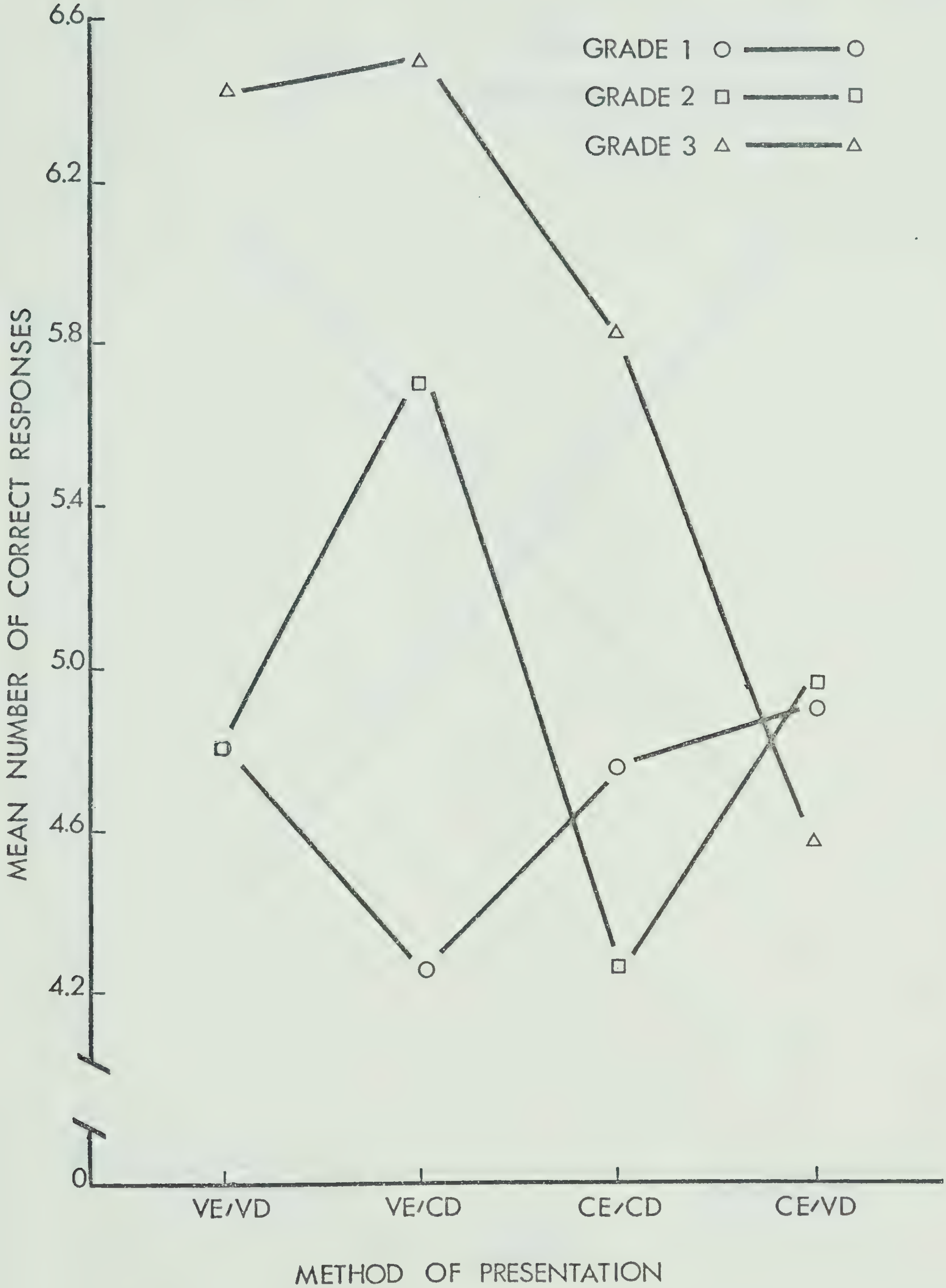

Figure 2: Age x Method of Presentation Interaction

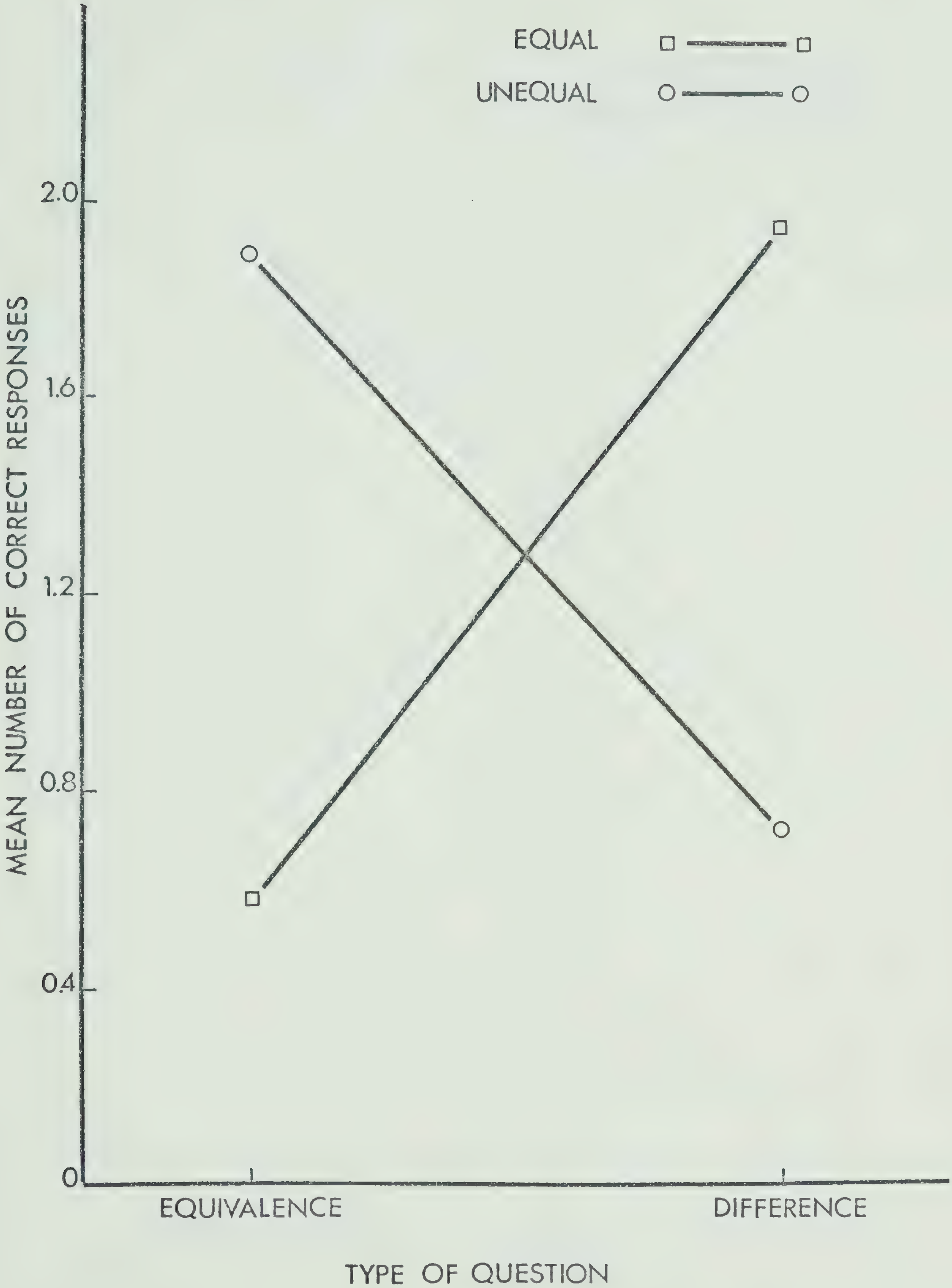

Figure 3: Type of Question x Stimulus Interaction: First Analysis

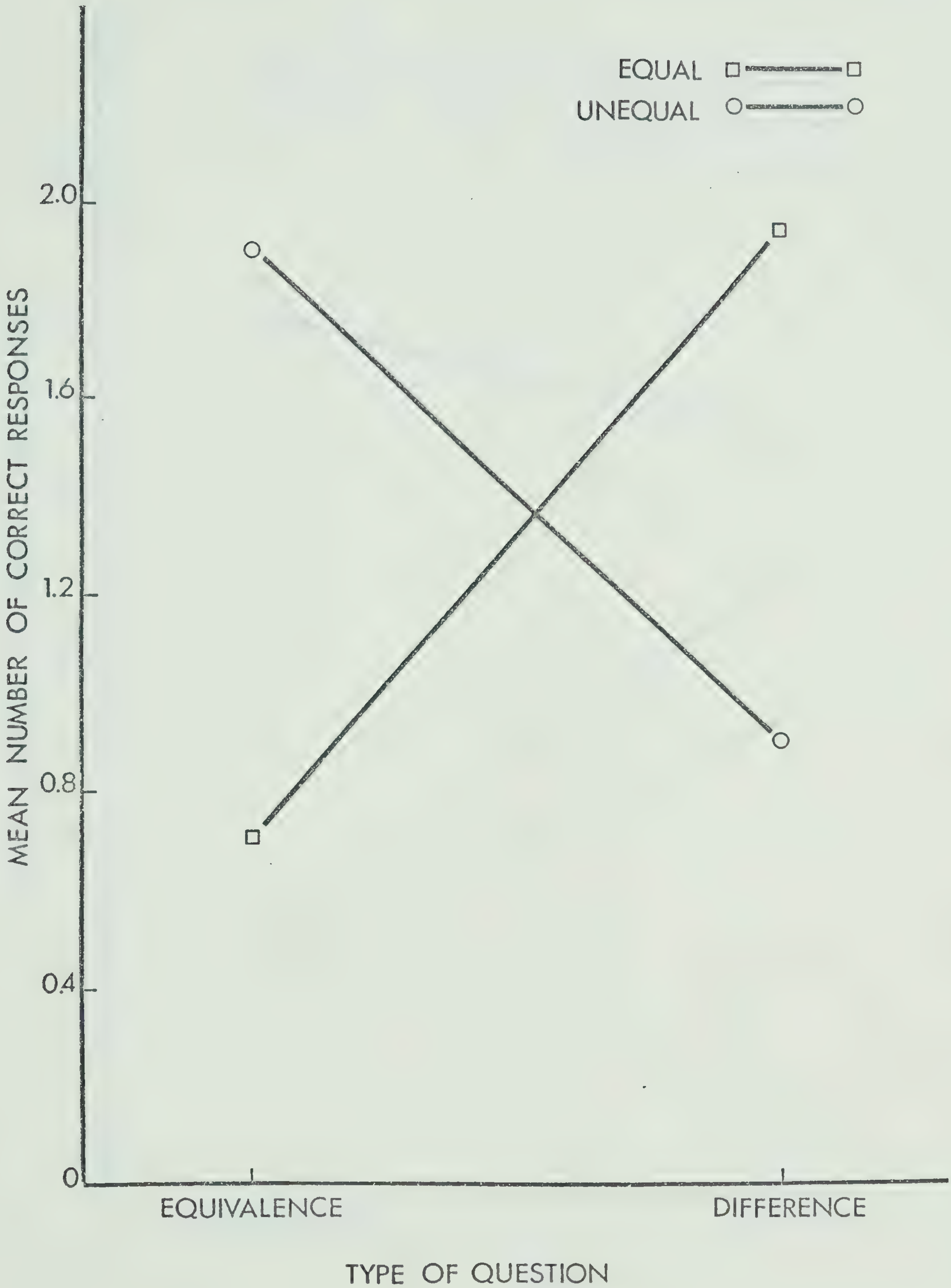

Figure 4: Type of Question x Stimulus Interaction: Second Analysis

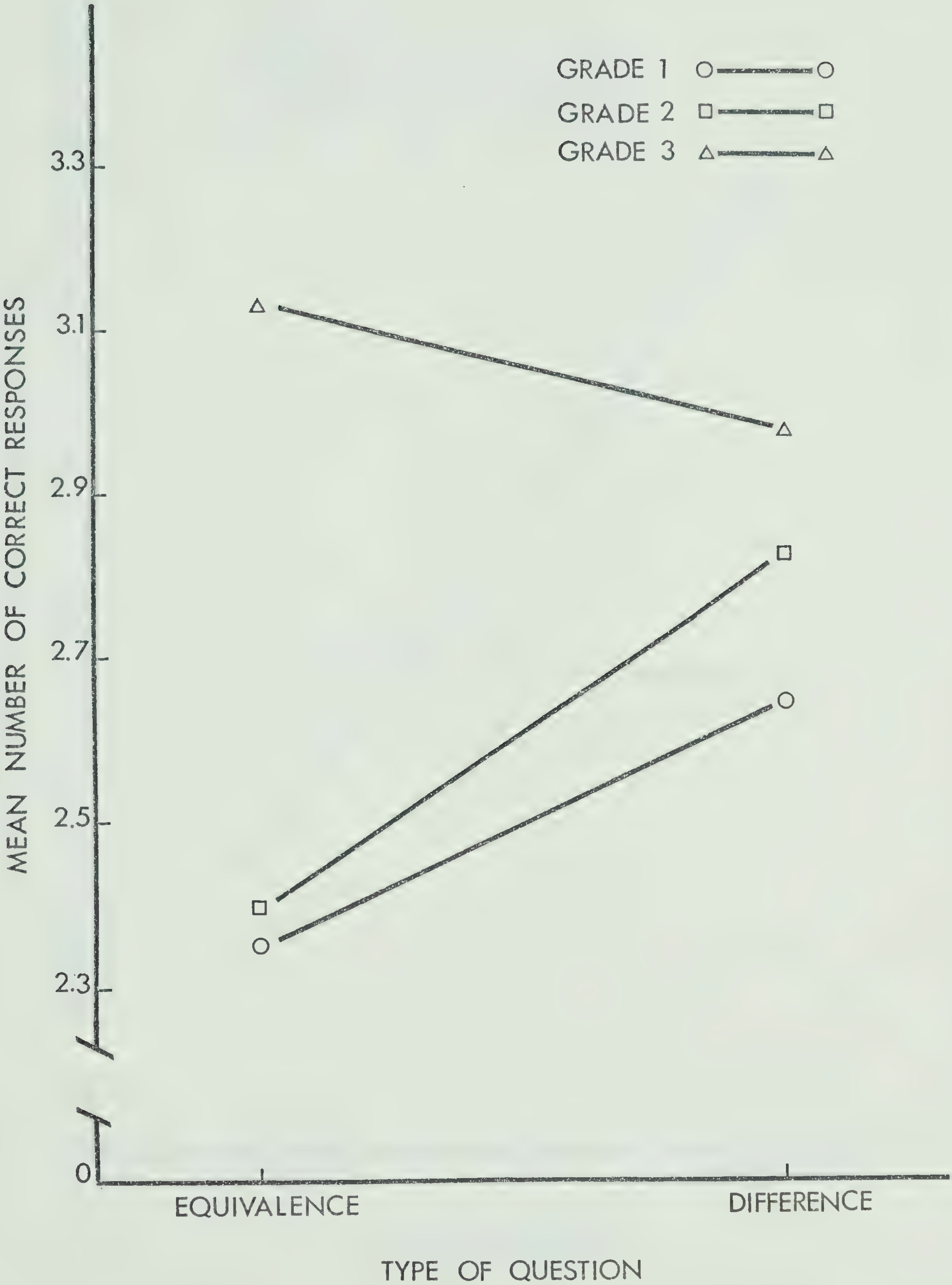

Figure 5: Type of Question x Age Interaction

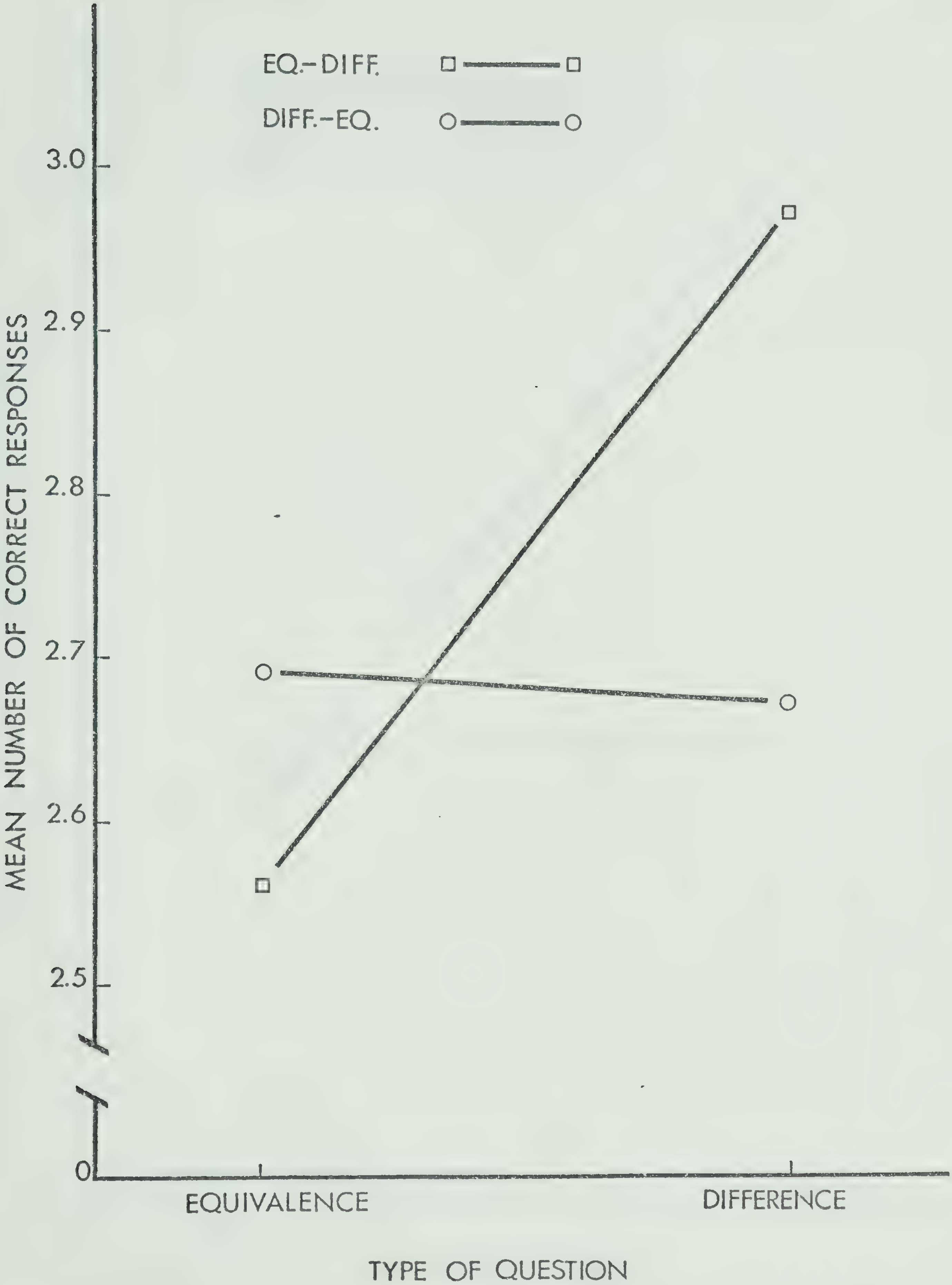

Figure 6: Graph of type of question x question order interaction based on means presented in Table 6.

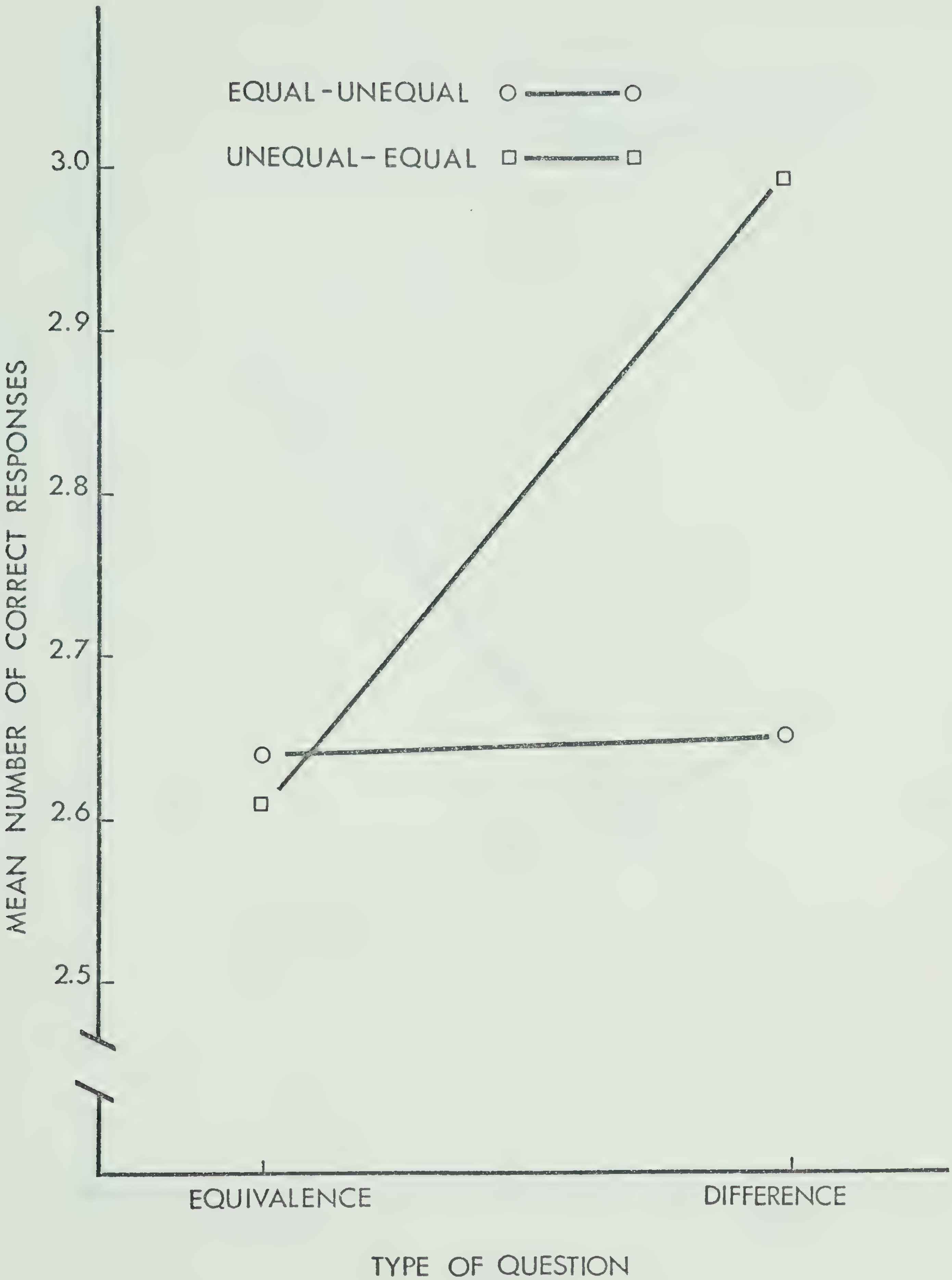

Figure 7: Graph of type of question x stimulus presentation order interaction based on means presented in Table 7.

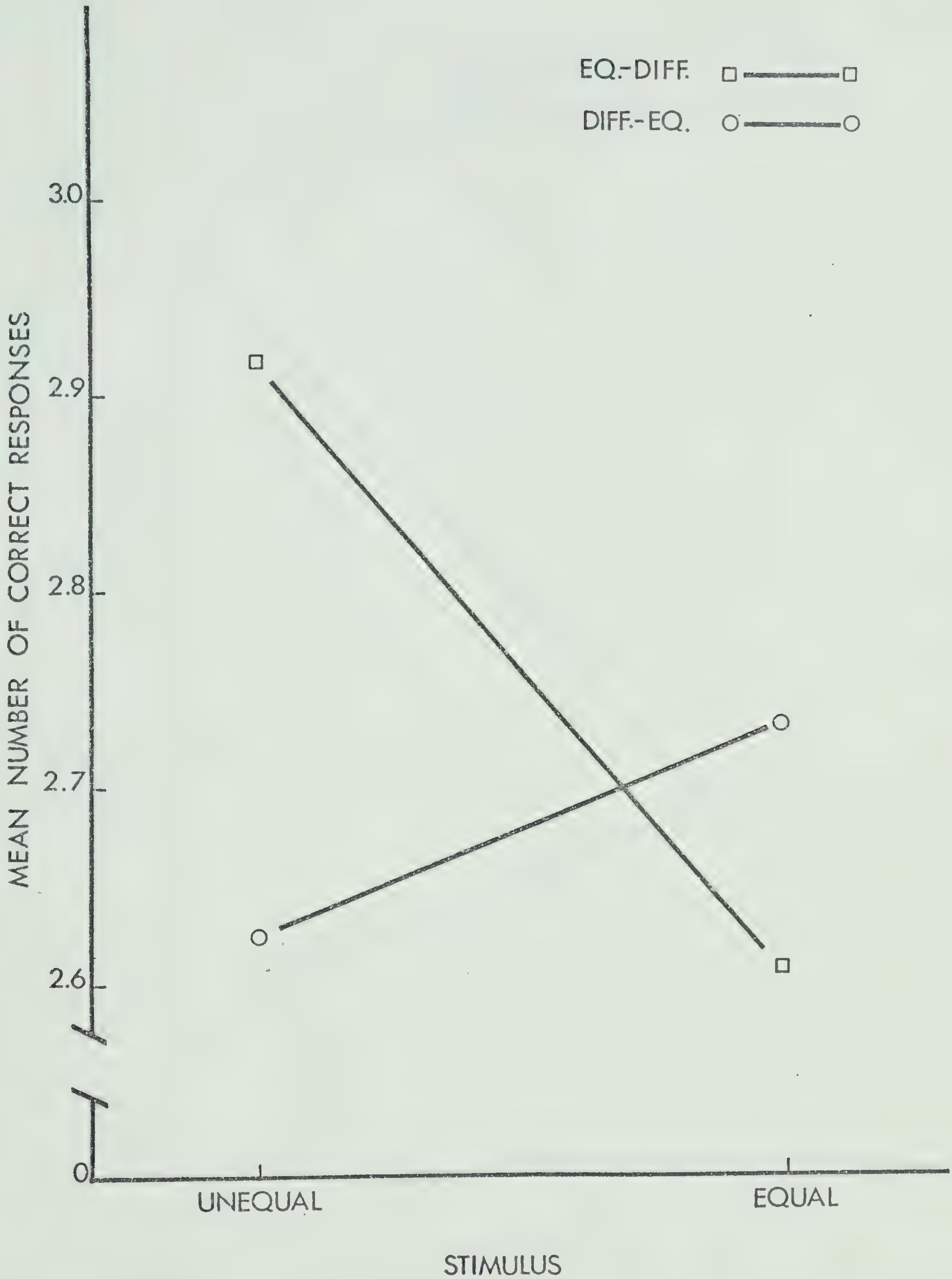

Figure 8: Graph of stimulus x question order interaction based on means presented in Table 8.

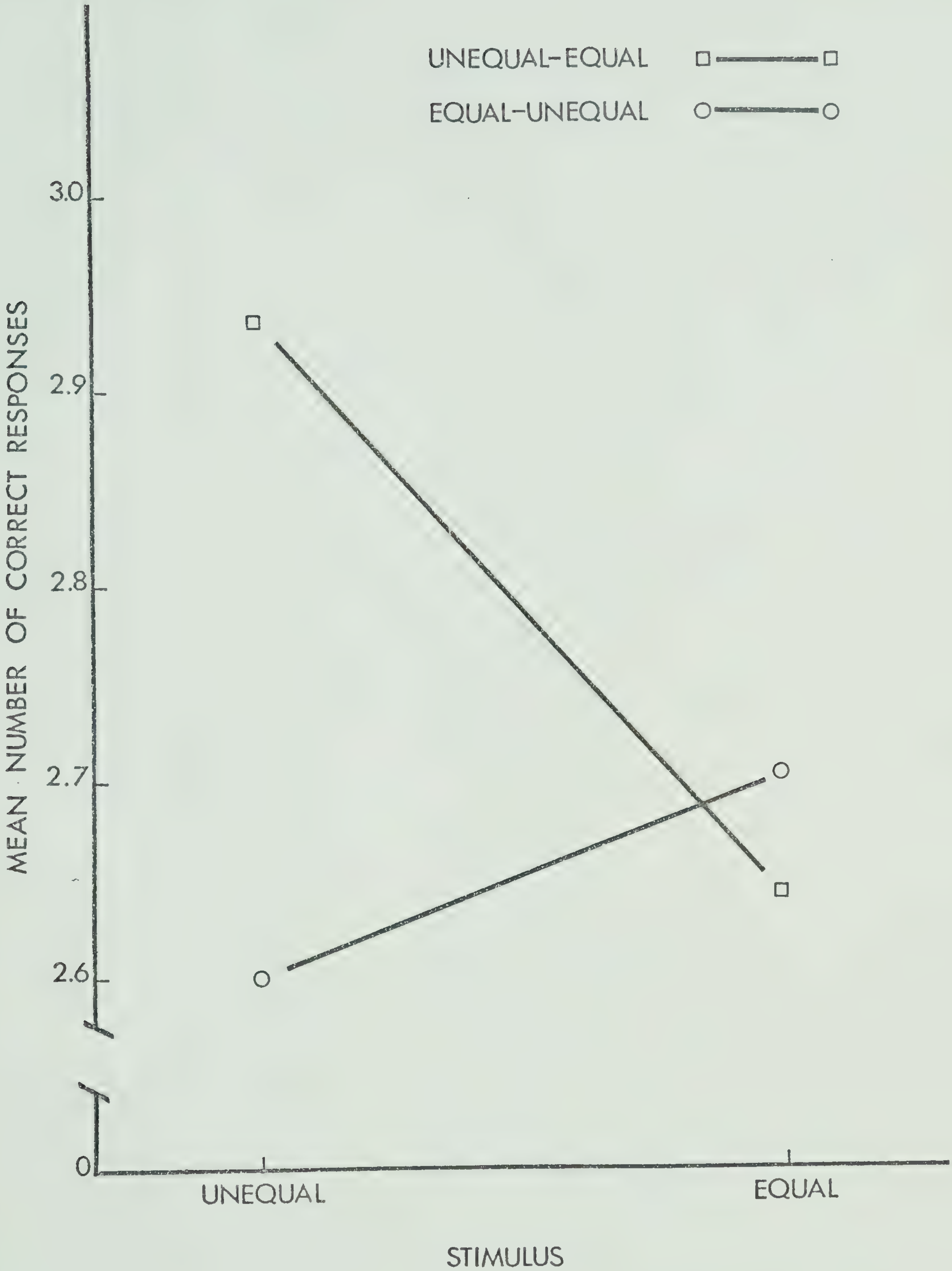

Figure 9: Graph of stimulus x stimulus presentation order interaction based on means presented in Table 9.